Chromosome level draft genomes of the fall armyworm, *Spodoptera frugiperda* (Lepidoptera: Noctuidae), an alien invasive pest in China

Huan Liu¹,¹⁰*, Tianming Lan¹,¹¹*, Dongming Fang¹,¹¹*, Furong Gui²,⁹*, Hongli Wang¹,¹⁰, Wei Guo³, Xiaofang Cheng⁴, Yue Chang⁵, Shuqi He², Lihua Lyu⁶, Sunil Kumar Sahu¹,¹⁰, Le Cheng⁷, Haimeng Li⁸, Ping Liu⁴, Guangyi Fan⁵, Tongxian Liu⁸, Ruoshi Hao⁹, Haorong Lu¹,¹⁰, Bin Chen⁹, Shusheng Zhu², Zhihui Lu⁹, Fangneng Huang¹², Wei Dong¹,¹⁰, Yang Dong², Le Kang³, Huanming Yang¹,¹⁰, Jun Sheng²,⁹*, Youyong Zhu²*, Xin Liu¹,¹⁰*.

¹. BGI-Shenzhen, Beishan Industrial Zone, Yantian District, Shenzhen 518083, China

². Yunnan Agricultural University, Kunming 650201, China

³. State Key Laboratory of Integrated Management of Pest Insects and Rodents, Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, China

⁴. MGI, BGI-Shenzhen, Shenzhen 518083, China

⁵. BGI-Qingdao, BGI-Shenzhen, Qingdao 266555, China

⁶. Plant protection research institute, Guangdong Academy of Agricultural Sciences, Guangzhou 510640, China

⁷. BGI-Yunnan, No. 389 Haiyuan Road, High-tech Development Zone, Kunming, Yunnan 650106, China

⁸. Plant Medical College, Qingdao Agricultural University. Qingdao 266109, China

⁹. Yunnan Plateau Characteristic Agriculture Industry Research Institute, Kunming 650201, China

¹⁰. CC-BY-NC 4.0 International license

¹¹. Certified by peer review is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.
Abstract:
The fall armyworm (FAW), Spodoptera frugiperda (J.E. Smith) is a severely destructive pest native to the Americas, but has now become an alien invasive pest in China, and causes significant economic loss. Therefore, in order to make effective management strategies, it is highly essential to understand genomic architecture and its genetic background. In this study, we assembled two chromosome scale genomes of the fall armyworm, representing one male and one female individual procured from Yunnan province of China. The genome sizes were identified as 542.42 Mb with N50 of 14.16 Mb, and 530.77 Mb with N50 of 14.89 Mb for the male and female FAW, respectively. We predicted about 22,201 genes in the male genome. We found the expansion of cytochrome P450 and glutathione s-transferase gene families, which are functionally related to the intensified detoxification and pesticides tolerance. Further population analyses of corn strain (C strain) and rice strain (R strain) revealed that the Chinese fall armyworm was most likely invaded from Africa. These strain information, genome features and possible invasion source described in this study will be extremely important for making effective strategies to manage the fall armyworms.

Key Words: Fall armyworm, Spodoptera frugiperda, Chromosome-level genome, Insect, Pest

1 Introduction
It has been more than 100 years since the Fall armyworm (FAW), Spodoptera frugiperda (J.E. Smith) was reported to damage maize and other crops in the USA. It is a severely destructive agricultural pest native to Americas which survives the whole year in the tropical and subtropical area from far south
Argentina, Chile and La Pampa to far north Florida, Texas, Mexico and the Caribbean. It cannot survive severe winters because of the lack of diapause.

However, FAW has a remarkable capacity of long-distance migration, with which the FAW can fly over 100 km per night. Each spring, it can migrate over 2000 km from the overwintering areas to reinvade more northern regions, even up to Canada. Recently, FAW spread out from its native region and invaded into Africa in 2016 with the report in São Tomé, Bénin, Togo and Nigeria. This invasion rapidly became widespread in the whole sub-Saharan Africa till October 2017. Following this trend, FAW soon invaded into many Asian countries, including India, Yemen, Thailand, Myanmar and Sri Lanka in 2018.

Early R et al (2018) forecasted that China is one of the most vulnerable countries of being invaded by the FAW according to the information on frequent commercial trade and passenger transportation between Africa and China. Half a year later after this forecast, in January 2019, International Plant Protection Convention (IPPC) Contact Point for China spotted FAW for the first time in Puer and Dehong city, Yunnan Province, China (https://www.ippc.int/en/news/first-detection-of-fall-armyworm-in-china/). This invasive pest has rapidly invaded many provinces in China by June 2019. Now, the FAW has been detected in large parts of the world (Figure 1).

There are several hosts of FAW, which mainly includes 186 plant species from 42 families. Although FAW is a highly polyphagous pest, graminaceous plants are their preferred hosts, such as maize, rice, sorghum and other major agricultural species. Especially, maize is most likely to be attacked than other plant species.

Maize is very important food security of many countries in Asia, Africa, and Latin America (https://maize.org/projects-cimmyt-and-iita-2/), and now are facing severe threats from the infestation of the FAW. The production of maize infested by the FAW can suffer yield loss of 40% to 72%, and in some plots 100% total has been reported. The yield loss of maize can reach 8.3 to 20.6m tons per year in just 12 African countries without any control methods for FAW,
according to the study conducted by Centre for Agriculture and Biosciences International (CABI). For Brazil alone, cost to control the FAW on maize is more than 600 million dollars per year. The economic and the yield losses by FAW is a major concern worldwide.

The FAW is not a new species to science; it has been an herbivorous pest for many years. However, the mostly common used approach to mitigate the damage to crops is still the broad-spectrum insecticides. The use of insecticides highly relies on the knowledge of farmers, but many farmers even do not know the name of the pest. Lack of scientific guidance leads to the inappropriate use of pesticides. Besides, the effective management of FAW may require five sprays of pesticides per maize cycle, and many smallholder farmers cannot afford the expensive costs for it, resulting in the use of low-quality products or low dose sprays. Moreover, the FAW always hides inside the stem of maize, this makes the insecticide much less effective. Although many pesticides are less harmful to the environment and humans, all these factors can lead the sublethal effects, which possibly help the FAW to evolve resistance against the pesticides. An effective compensating management for insecticides is the use of *Bacillus thuringiensis* (Bt) toxins produced by the bacterium. Bt plants have been proved fatal to many insect pests, including the FAW. The Bt toxin provides much longer protection than insecticides and less harmful to the environment and humans. Although some research reported the resistance of FAW to Bt maize, multiple genes or new gene with more or new Bt toxin expressed are still thought to have a good performance for resisting the FAW. Biological control, including the introduction of natural enemies and using companion cropping system, is also an effective way to resist the FAW.

The FAW at least consist of two morphological identical but genetically distinct subpopulations, the corn strain (C strain) and rice strain (R strain). The two strains have their own preferable host plants. The C strain is preferentially associated with maize and other large grasses, but the C strain prefers rice and large grasses. The C strain is subdivided into two subgroups, the FL-type and the TX-type. The TX-type is distributed in most of the Americas, but the FL-type is only limited to Florida and the Caribbean. Each strain has its strain-specific physiological traits, leading to some strain-specific response to biological and chemical agents. The C strain larvae are more tolerant than the R strain to the methyl parathion, cypermethrin, cypermethrin and δ-endotoxin from transgenic Bt plant. Therefore, the origin of FAW needs to be considered to make effective strategies to manage the FAW. The *Cytochrome c oxidase subunit I* (COI) and the *Triosephosphate isomerase* (Tpi) gene are also selected for identifying
the subtype of the FAW, but these markers cannot always give the right identifications\textsuperscript{44}.

The management of the FAW needs more detailed genetic information to help people know more about the FAW, to find more new genes to develop more effective Bt plants, and to more accurately identify the different strains for precision spraying of pesticides. Although several genomes of the FAW has been sequenced and assembled\textsuperscript{45-47}, the assemblies were fragmented. Moreover, two genomes were assembled using the Sf21\textsuperscript{45} and Sf9\textsuperscript{47} cell line, respectively, the resource was unique that cannot well provide a comprehensive reference. In this study, we assembled two FAW genomes to the chromosome level using two samples (one male, one female) collected from Yunnan Province, China. We analyzed the subtype of the FAW invaded Yunnan province, and also discussed the possible resource of the invaded FAW in China. We are also screening expanded gene families to seek some key genes with the function of polyphagia and tolerance to insecticides. China is the second largest corn producer after the USA; therefore, it is urgent to select a series of methods to control the FAW. This study provided key information to help make strategies to manage the FAW in China.

2 Materials and Methods

2.1 Samples and treatments

We collected seven FAW samples, including four adults, two fifth-instar larvae and two sixth-instar larvae (Table 1). The four adult FAWs were collected from Yunnan Province, China, and the larvae were collected from Guangdong Province, China. One male and one female adult individual were used for genome sequencing and assembly (Figure 2). Two other adult FAWs were used to capture the conformation of chromosomes to perform the chromosome level genome assembly. One fifth-instar larva and one sixth-instar larva were subjected for the transcriptomic studies. One sixth-instar larva was used for whole genome sequencing with 5K and 300bp insert size. All samples were intestinal and ovarian, and thoroughly cleaned before performing DNA or RNA isolation.

<table>
<thead>
<tr>
<th>Sample Identifier</th>
<th>Species</th>
<th>Location</th>
<th>Sex</th>
<th>Instars</th>
</tr>
</thead>
<tbody>
<tr>
<td>SFynMstLFR</td>
<td>Spodoptera frugiperda</td>
<td>Kunming, Yunnan</td>
<td>Male</td>
<td>Adult</td>
</tr>
<tr>
<td>SFynFMstLFR</td>
<td>Spodoptera frugiperda</td>
<td>Kunming, Yunnan</td>
<td>Female</td>
<td>Adult</td>
</tr>
<tr>
<td>SFgdRNA 1</td>
<td>Spodoptera frugiperda</td>
<td>Guangzhou, Guangdong</td>
<td>Unknown</td>
<td>Fifth-instar</td>
</tr>
<tr>
<td>SFgdRNA 2</td>
<td>Spodoptera frugiperda</td>
<td>Guangzhou, Guangdong</td>
<td>Unknown</td>
<td>Sixth-instar</td>
</tr>
<tr>
<td>SFgdWGS</td>
<td>Spodoptera frugiperda</td>
<td>Guangzhou, Guangdong</td>
<td>Unknown</td>
<td>Sixth-instar</td>
</tr>
</tbody>
</table>
Spodoptera frugiperda
Kunming, Yunnan
Unknown
Adult

Figure 2 The two adult fall armyworm (FAWs) used for genome assembly. The left FAW is female, the right FAW is the male individual.

2.2 DNA isolation, library preparation, sequencing and genome assembly.

The high molecular weight DNA was extracted using the separated muscle tissue by following the protocol recommended by Cheng et al (2018)\textsuperscript{48}. We then used the single tube long fragment read (stLFR) technology\textsuperscript{49} to preparing the co-barcoding DNA libraries with the MGIEasy stLFR Library Prep Kit (lot number: 1000005622), and the libraries were then loaded on the sequencer for sequencing according to the protocol of MGISEQ-2000\textsuperscript{50}. To testify the accuracy of genome assembly, we constructed a DNA library with a 5kb insert size and sequenced by BGISEQ-500 sequencer. To finally ligate the scaffolds to chromosomes, Hi-C technology\textsuperscript{51} was used to capture the conformation of chromosomes using another two adult individuals. The primary genome was assembled using the supernova (v2.1.1)\textsuperscript{52} software with the parameters --maxreads=700M. We filled gaps by use of GapCloser\textsuperscript{53} and GapCloser stLFR (unpublished method) with the default parametes. Finally, we performed the chromosome concatenation using the Hi-C generated data by 3d-DNA pipeline\textsuperscript{54}.

2.3 Comparative genomics analysis
Identification of orthology and paralogy groups of *Spodoptera frugiperda* genes and other considered genomes were done using OrthoMCL\textsuperscript{55} methods on the all-versus-all BLASTP alignment (e-value, <1e-5). We constructed gene families for nine species including *Bombyx mori*, *Danaus plexippus*, *Drosophila melanogaster*, *Heliconius melpomene*, *Helicoverpa armigera*, *Manduca sexta*, *Plutella xylostella*, *Spodoptera litura* and *Spodoptera frugiperda*. The phylogenetic tree, including these nine species, was constructed using the combined set of all the single copy genes.

To search for homology, we compared protein-coding genes of *Spodoptera frugiperda* to that of other species using BLASTP with an E-value threshold of 1e-5. Based on the Whole-genome BLASTP and the genome annotation results, we detected the syntenic blocks using MCscan\textsuperscript{56}. A region with at least five syntenic genes and no more than 15 gapped genes was defined as a syntenic block.

2.4 RNA isolation, transcriptome libraries preparation and sequencing

The RNA extraction kit (RNeasy Mini Kit, Qiagen) was used for the total RNA isolation. We performed the RNase-free agarose gel to check the contamination, and then the RNA integrity and purity were measured by Agilent 2100 Bioanalyzer system (Agilent, United States) and NanoDrop Spectrophotometer (THERMO, United States), respectively. The extracted RNA was fragmented into 200-400 bp and reverse transcribed to cDNA for library preparation. The libraries were prepared to follow the manufacturer’s instructions for the BGISEQ-500 sequencing platform. Pair-end 100 sequencing was performed on the BGISEQ-500 sequencer using the processed libraries.

2.4 Bioinformatics analysis for transcriptome data

Raw data were firstly processed using the Trimmomatic to filter the reads with adaptors and reads with the proportion of Ns and low-quality bases larger than 10\% by SOAPfilter. Bridger software (v20141201) was used to *de novo* assemble the transcriptome, the reluctances were then removed by TGICL. The contigs were concatenated into scaffolds and further assembled to unigenes by clustering and removing redundancy.

FPKM was calculated to estimate the expression level of unigenes. In the study, the reads were mapped against the unigene library using Bowtie, and then unique mapped reads were selected for estimating the expression level by combining eXpress. Finally, DEG unigenes were selected with differential expression level with the parameter of FDR ≤ 0.01 and Fold Change ≥4.

2.5 Identifying the strains and the possible source of FAW invasion in China
We used the Tpi gene as a DNA maker to identify the strain of the FAW invaded into China. We identified the Tpi gene fragments from four FAWs, including two from Yunnan province (sequences were retrieved from the whole genome sequencing data) and two from Guangdong province (sequences were retrieved from the RNA-seq data). Eight sites (TpiE4-129, TpiE4-144, TpiE4-165, TpiE4-168, TpiE4-180, TpiE4-183, TpiE4-192, TpiE4-198) in the fourth exon of the Tpi gene were used for determining the strains and a possible source of the invaded FAW\textsuperscript{57}. The fourth intron of the Tpi gene\textsuperscript{57} was used for constructing the phylogenetic tree using the PhyML (v3.0)\textsuperscript{58} software with the Maximum-Likelihood methods to assist in identifying the strains. Only two Tpi fourth introns were used for phylogenetic analysis because we cannot retrieve the introns from RNAseq data. 112 sequences of the fourth intron of the Tpi gene with strain information were downloaded from NCBI (Table S1). The sequence alignments were performed using the Clustal W\textsuperscript{59}.

3 Results and discussion

3.1 Chromosome level genome assembly for two FAWs

Only one individual was used for genome sequencing for both the male and female genome assembly, which was less than the number used for genome assembly by Gouin A et al (2017)\textsuperscript{46}. This maximally decreased the heterozygosity level. A total of 2μg and 1.6μg total genomic DNA with the average band size larger than 20 kb were isolated for stLFR libraries preparation from the male and female FAW, respectively. Two stLFR libraries with more than 30 million barcodes were constructed for running on the BGISEQ-500 sequencer. 110.93 Gb and 91.41 Gb high-quality reads were generated for the male and female individual, respectively (Table S2). The primary assembly sizes were 542.42 Mb and 530.77 Mb for the male and female individual. The scaffold N50 and N90 for the male individual were 507.12 Kb and 6.43 Kb, and that for the female were 528.27 Kb and 5.11 Kb, respectively (Table 2). After getting the confirmation from Hi-C sequencing, we finally concatenated the scaffolds to 31 chromosomes with the scaffold N50 of 14.16 Mb and 14.89 Mb for the male and female individual, respectively. Genomes sizes assembled in this study are in the range of the lepidoptera genome sizes at 246M to 809M\textsuperscript{60}, but are larger than previously assembled FAW genomes\textsuperscript{45-47}, probably due to the use of the new stLFR technology with the super high coverage and reads quality. This is the first time to assemble the genome of the FAW to the chromosome level, which will be in no doubt to accelerate the biological studies and making the effective strategy of pest managements.
3.2 Evaluation of the assembly

The GC content for the genomes of the male and female individual was found to be 35.52% and 36.57% (Table 2), respectively, which showed a similar level with most closely related Lepidoptera species ranging from 31.6% to 37.7%\(^1\). We used the Benchmarking Universal Single-Copy Orthologs (BUSCO: version 2.0)\(^2\) and Core Eukaryotic Genes (CEGMA)\(^3\) to evaluated the completeness of the two assemblies. In BUSCO analysis, genomes of the male and female samples covered 95% and 94.5% complete BUSCO genes (Table S3). In the CEGMA analysis, 83.47% and 85.48% complete core eukaryotic genes were found for the two genomes (Table S4). This is better than all FAW genomes that has been published (PRJNA380964; PRJNA257248; PRJEB13110; and PRJNA344686). Besides, we also mapped the sequencing data generated from libraries of Hi-C, MatePair5K, WGS, and the RNA-seq to the assembled male genome. The mapping rates were all higher than 90% (Table 3), and the insert size were also consistent with the libraries, except for the MatePair5K, probably because the large insert size cannot ensure that the one pair reads they mapped to the same scaffold. EST sequences of the FAW were downloaded from NCBI and the transcripts were assembled without reference. We mapped these EST sequences and transcripts to the male reference genome we assembled, and the results showed that more than 90% EST sequences and more than 80% transcripts we assembled could be found on the assembled male genome (Table S5). However, it is noteworthy that the transcript from SFgdRNA 2 has a lower mapping rate than that of SFgdRNA 1. We inferred that this resulted due to the genetic differences between the C strain and the R strain, because the SFgdRNA 2
sample was identified belong to the R strain. Overall, all the above results well testified the completeness of the two genomes.

### Table 3 Mapping reads against the assembled male genome using raw reads generated by different libraries

<table>
<thead>
<tr>
<th>Type</th>
<th>HIC</th>
<th>MatePair</th>
<th>stLFR</th>
<th>RNA-seq</th>
<th>WGS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Mapped Reads</td>
<td>93.58%</td>
<td>93.68%</td>
<td>95.60%</td>
<td>98.98%</td>
<td>90.71%</td>
</tr>
<tr>
<td>Perfect Match</td>
<td>38.81%</td>
<td>38.54%</td>
<td>43.67%</td>
<td>52.14%</td>
<td>27.76%</td>
</tr>
<tr>
<td>Unique Match</td>
<td>76.39%</td>
<td>77.15%</td>
<td>83.23%</td>
<td>76.99%</td>
<td>74.77%</td>
</tr>
<tr>
<td>Total Unmapped Reads</td>
<td>6.42%</td>
<td>6.32%</td>
<td>4.40%</td>
<td>1.02%</td>
<td>9.29%</td>
</tr>
<tr>
<td>Total FullMapped Reads</td>
<td>32.48%</td>
<td>30.06%</td>
<td>71.45%</td>
<td>52.46%</td>
<td>49.40%</td>
</tr>
</tbody>
</table>

3.3 Annotation

We firstly used Repeat Modeler (v1.0.11), LTR finder (v1.0.5) and repeatscount (v1.0.5) methods to identify de novo repeat motifs by modeling ab initio, and these repeat motifs were added into the RepBase library as known repeat elements. We then performed the RepeatMasker library to mask the assembly, using the combined RepBase library. Usually, Repeat elements take a substantial part of the genome and contribute as important events to genome evolution. In this study, by the combination of de novo and homology-based searching, 153 Mbp repeat elements were finally identified for the male FAW, and accounting for 28.24% the FAW genomes.

Gene prediction was carried out by both the homology-based and de novo methods using repeat masked genomes. For the de novo prediction, we used Augustus, glimmerHMM and SNAP (Table S6). For the homology-based approaches, *Bombyx mori*, *Danaus plexippus*, *Drosophila melanogaster* and *Spodoptera litura* genomes were used for homology alignments using the TblastN. Moreover, transcripts that were predicted with RNA-seq, Gene sets were then merged to form a non-redundant gene set with GLEAN; then all annotated genes were checked and filtered manually. A total of 22201 genes were finally obtained for the male samples (Table S6).

In the final gene set we identified, we found 94.2% compete for BUSCO genes and 95.16% CEGMA genes, which were all better than the published FAW genome (Table S7, Table S8). Of these identified genes, 93.48% was confirmed that have functions (Table S9), which was facilitated the further exploration of the functions. Besides, we also found 60 miRNAs, 840 tRNAs and 197 rRNAs by using the homology prediction method.

3.4 The transcriptome analysis of the larvae
After filtering, we finally obtained 58Gb clean data with 341,526,489 cleaned reads. These reads were assembled into 72,604 contigs with the N50 of 2077bp. These contigs were further assembled into scaffolds, and the scaffolds were further assembled to 51,495 unigenes by clustering and removing redundancy. The contig number in our study are significantly higher than that in the study of Kakumani et al. This maybe result from that we only used a single method for assembly. We also calculated the expression abundance for unigenes between the fifth-instar and sixth-instar larvae. The result showed 2,648 differentially expressed genes (DEGs). We further performed the clustering analysis to cluster genes with identical or similar expressed behaviors. Remarkable expression difference was found between the fifth-instar larvae and the sixth-instar larvae (Figure S1). This difference was in consistent with the different strains of the two larvae (we described in 3.6). However, if the different instar contributes to the differential expression, it was further confirmed by more detailed analysis.

3.5 Comparison to other published lepidopteran genomes

To further explore the detailed relationship between the FAW and its other lepidopteran relatives. We constructed a phylogenetic tree of nine genomes using 2,001 single copy genes downloaded from NCBI and insectbase (Table S10). The result showed that the S. frugiperda which we sequenced actually clustered with its most related species S. litura, which is in accordance with the study by Cheng et al. (Figure 3).

Figure 3 The phylogenetic relationships among nine lepidopteran genomes.

Through the gene family analysis, a total of 12,516 gene families were found in the S. frugiperda genome, including 20,012 genes. Of these gene families, 324 are specific to the S. frugiperda compared to the other seven species (Table S11). Then, we analyzed the 34 functional gene families of insects, finding some expanded gene families, including cytochrome p450, glutathione s-transferase, and hydrolase (Table 4). The cytochrome p450 gene family is closely related with
intensified detoxification\textsuperscript{69}, the genes in this family of the \textit{S. frugiperda} is 200, more than that of \textit{S. litura}, which indicated that the \textit{S. frugiperda} was more polyphagous than \textit{S. litura}. This is also consistent with the habits of \textit{S. frugiperda}. The expanded glutathione s-transferase gene family was proved that could enhance the insecticides tolerance of the \textit{S. litura}\textsuperscript{69}. In this study, we found more genes for \textit{S. frugiperda}, which indicated that the \textit{S. frugiperda} was probably easier to gain resistance to pesticides. These gene families are a valuable genetic source to develop more effective pesticides or other methods to manage the FAW.

### Table 4 Identified genes in gene families of 9 insects

<table>
<thead>
<tr>
<th>Gene Family</th>
<th>DME</th>
<th>PXY</th>
<th>MSE</th>
<th>BMO</th>
<th>HME</th>
<th>DPL</th>
<th>HAR</th>
<th>SLI</th>
<th>SFR</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABC transporter</td>
<td>55</td>
<td>76</td>
<td>40</td>
<td>40</td>
<td>35</td>
<td>38</td>
<td>36</td>
<td>86</td>
<td>66</td>
</tr>
<tr>
<td>acetylcholine receptor</td>
<td>18</td>
<td>27</td>
<td>20</td>
<td>27</td>
<td>24</td>
<td>28</td>
<td>22</td>
<td>27</td>
<td></td>
</tr>
<tr>
<td>Acetylcholinesterase</td>
<td>27</td>
<td>39</td>
<td>62</td>
<td>51</td>
<td>32</td>
<td>27</td>
<td>82</td>
<td>87</td>
<td>46</td>
</tr>
<tr>
<td>alkaline phosphatase</td>
<td>13</td>
<td>10</td>
<td>11</td>
<td>10</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>11</td>
<td>12</td>
</tr>
<tr>
<td>aminopeptidase</td>
<td>45</td>
<td>53</td>
<td>49</td>
<td>37</td>
<td>35</td>
<td>39</td>
<td>15</td>
<td>19</td>
<td>24</td>
</tr>
<tr>
<td>carboxylesterase</td>
<td>30</td>
<td>55</td>
<td>90</td>
<td>70</td>
<td>50</td>
<td>54</td>
<td>100</td>
<td>110</td>
<td>84</td>
</tr>
<tr>
<td>Chitinase</td>
<td>45</td>
<td>27</td>
<td>28</td>
<td>24</td>
<td>26</td>
<td>32</td>
<td>23</td>
<td>26</td>
<td>46</td>
</tr>
<tr>
<td>chloride channel</td>
<td>16</td>
<td>25</td>
<td>20</td>
<td>23</td>
<td>18</td>
<td>19</td>
<td>19</td>
<td>16</td>
<td>24</td>
</tr>
<tr>
<td>CTL</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>cytochrome p450</td>
<td>104</td>
<td>108</td>
<td>144</td>
<td>96</td>
<td>130</td>
<td>99</td>
<td>121</td>
<td>132</td>
<td>200</td>
</tr>
<tr>
<td>DNA methyltransferase</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>ecdysone receptor</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>GABA</td>
<td>15</td>
<td>16</td>
<td>32</td>
<td>9</td>
<td>12</td>
<td>15</td>
<td>16</td>
<td>14</td>
<td>10</td>
</tr>
<tr>
<td>glutamate-gated chloride channel</td>
<td>10</td>
<td>14</td>
<td>12</td>
<td>13</td>
<td>10</td>
<td>10</td>
<td>12</td>
<td>12</td>
<td>14</td>
</tr>
<tr>
<td>Glutathione s-transferase</td>
<td>50</td>
<td>28</td>
<td>42</td>
<td>25</td>
<td>23</td>
<td>25</td>
<td>46</td>
<td>47</td>
<td>60</td>
</tr>
<tr>
<td>glycosyltransferase</td>
<td>65</td>
<td>25</td>
<td>23</td>
<td>19</td>
<td>24</td>
<td>22</td>
<td>19</td>
<td>21</td>
<td>31</td>
</tr>
<tr>
<td>G protein</td>
<td>23</td>
<td>42</td>
<td>36</td>
<td>31</td>
<td>26</td>
<td>29</td>
<td>29</td>
<td>29</td>
<td>30</td>
</tr>
<tr>
<td>gustatory receptor</td>
<td>54</td>
<td>69</td>
<td>45</td>
<td>76</td>
<td>73</td>
<td>68</td>
<td>197</td>
<td>238</td>
<td>220</td>
</tr>
<tr>
<td>heat shock protein</td>
<td>45</td>
<td>38</td>
<td>54</td>
<td>35</td>
<td>29</td>
<td>46</td>
<td>40</td>
<td>48</td>
<td>45</td>
</tr>
<tr>
<td>hydrolase</td>
<td>213</td>
<td>250</td>
<td>243</td>
<td>209</td>
<td>181</td>
<td>204</td>
<td>267</td>
<td>310</td>
<td>396</td>
</tr>
<tr>
<td>immunoglobulin</td>
<td>65</td>
<td>51</td>
<td>70</td>
<td>76</td>
<td>56</td>
<td>69</td>
<td>71</td>
<td>67</td>
<td>80</td>
</tr>
<tr>
<td>odorant-binding protein</td>
<td>62</td>
<td>65</td>
<td>70</td>
<td>51</td>
<td>87</td>
<td>63</td>
<td>77</td>
<td>61</td>
<td>70</td>
</tr>
<tr>
<td>odorant receptor</td>
<td>67</td>
<td>82</td>
<td>79</td>
<td>43</td>
<td>73</td>
<td>64</td>
<td>87</td>
<td>75</td>
<td>75</td>
</tr>
<tr>
<td>Painless</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>pheromone</td>
<td>43</td>
<td>22</td>
<td>29</td>
<td>15</td>
<td>19</td>
<td>14</td>
<td>16</td>
<td>14</td>
<td>19</td>
</tr>
<tr>
<td>protease inhibitor</td>
<td>75</td>
<td>33</td>
<td>60</td>
<td>38</td>
<td>26</td>
<td>36</td>
<td>46</td>
<td>39</td>
<td>51</td>
</tr>
<tr>
<td>Ryanodine receptor</td>
<td>6</td>
<td>12</td>
<td>3</td>
<td>7</td>
<td>4</td>
<td>8</td>
<td>6</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>sensory neuron membrane protein</td>
<td>3</td>
<td>5</td>
<td>4</td>
<td>5</td>
<td>4</td>
<td>7</td>
<td>7</td>
<td>3</td>
<td>10</td>
</tr>
<tr>
<td>serpin</td>
<td>35</td>
<td>23</td>
<td>30</td>
<td>18</td>
<td>19</td>
<td>24</td>
<td>21</td>
<td>18</td>
<td>26</td>
</tr>
<tr>
<td>sirtuin</td>
<td>6</td>
<td>4</td>
<td>3</td>
<td>5</td>
<td>5</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>7</td>
</tr>
</tbody>
</table>
3.6 FAW in China includes both the C and R strain, possibly invaded from Africa.

There are three strain-specific sites (E4165, E4168 and E4183) in the fourth exon of the Tpi gene that can identify the C strain from the R strain for Western Hemisphere populations. Especially, the E4183 is an effective diagnostic marker for Tpi gene to define the C or R strain. In this study, two Yunnan samples and one Guangdong sample were identified as the C strain and the other one Guangdong sample was identified as the R strain (Figure 4). The phylogenetic tree showed that the two Yunnan samples were clustered in the clade that consisted of all C strain individuals, which strengthened the results inferred using the strain-specific sites in the fourth exon (Figure 5). This result at least showed that the FAW invaded into China included both the strains. However, the detailed population genetic structure and the frequencies of the two strains in the Chinese population need more information from the population-level studies.
To further confirm the possible source of the Chinese FAW, we compared the haplotype consisted of all eight polymorphisms in the fourth exon of the Tpi gene as shown in the Figure 4. We found that all the four individuals hold identical haplotype with the African population, including three TpiC1 and one TpiRa1. Although the haplotype of C strain was shared by the African and the Western Hemisphere populations57, the TpiRa1 has not been detected in any Western Hemisphere populations57 showing the uniqueness to African populations. The finding of the TpiRa1 haplotype in the Guangdong population indicated that there were at least parts of the FAW populations in China that was invaded from Africa, probably through the frequent commercial trade and passenger transportation between Africa and China5. However, we cannot confirm other sources because of the small sample size we used here. The strain information and possible invasion source found in this study will be extremely important for making effective strategies to manage the FAWs in China.

Figure 5 The phylogenetic tree to identify the FAW strains of collected from Yunnan, China. The red branched presented the C strain, the red triangles present the two Yunnan samples.

4 Conclusions
In summary, we assembled two chromosome scale genomes of the fall armyworm, representing one male and one female individual procured from Yunnan province of China. The genome sizes were identified as 542.42 Mb with N50 of 14.16 Mb, and 530.77 Mb with N50 of 14.89 Mb for the male and female FAW, respectively. The completeness of the two genomes are better than all
previously published FAW genomes which is evident by the BUSCO and
CEGMA analysis. A total of 22,201 genes were predicted in the male genome,
and 12,516 gene families were found in the *S. frugiperda* genome, including 20,012
genes. Of these gene families, we found expansion of cytochrome p450 and
glutathione s-transferase gene families, which were closely related to the function
of intensified detoxification and pesticides tolerance. We finally identified both
the R strain and C strain individuals in the Chinese population, showed that the
Chinese FAW was most likely invaded from Africa. The strain information and
possible invasion source found in this study will be extremely important for
making the effective strategies to manage the FAWs in China.

### Acknowledgements

This study was supported by the Guangdong Provincial Key Laboratory of core
collection of corp genetic resources research and application
(NO.2011A091000047), Guangdong Provincial Key Laboratory of Genome Read
and Write (grant no. 2017B030301011) and Key Laboratory of Genomics, Ministry
of Agriculture.

### Data availability

The Raw sequencing data and the two chromosome level genome assemblies
have been deposited to the CNSA (CNGB Nucleotide Sequence Archive) with
accession CNP0000513 (https://db.cngb.org/cnsa/).

### Author Contributions

Huanming Yang, Le Kang, Jun Sheng, Youyong Zhu, Yang Dong, Xin Liu and
Huan Liu designed the research. Dongming Fang, Tianming Lan and Yue Chang,
Hongli Wang, Fangneng Huang, Wei Dong and Guangyi Fan performed the data
analysis. Xiaofang Chen, Haorong Lu, Ping Liu, Tongxian Liu, Rushi Hao, Bin
Chen, Shusheng Zhu, Zhihui Lu and Haimeng Li performed the DNA and RNA
extraction and the library preparation. Huan Liu, Tianming Lan, Yang Dong, Wei
Guo, Shuqi He, Le Chen and Lihua Lyu collected the samples. Tianming Lan,
Dongming Fang, Hongli Wang, Sunil Kumar Sahu and Furong Gui wrote and
revised the manuscript. All the authors read and revised the final version of the
manuscript.

### Competing interests

The authors declare no competing interests.

---

1 Hinds, W. E. & Dew, J. A. The grass worm or fall army worm. *Alabama
Agricultural and Experiment Station Bulletin* **186**, 25 (1915).


15 CABI. Invasive Species Compendium. [https://www.cabi.org/isc/fallarmyworm](https://www.cabi.org/isc/fallarmyworm).


29 Bernardi, D. *et al.* Cross-Resistance between Cry1 Proteins in Fall Armyworm (Spodoptera frugiperda) May Affect the Durability of Current Pyramided Bt


Nagoshi, R. N. Evidence that a major subpopulation of fall armyworm found in the Western Hemisphere is rare or absent in Africa, which may limit the range of crops at risk of infestation. *PloS one* **14**, e0208966, doi:10.1371/journal.pone.0208966 (2019).


Nandakumar, S., Ma, H. & Khan, A. S. Whole-Genome Sequence of the Spodoptera frugiperda Sf9 Insect Cell Line. *Genome Announcements* **5**:e00829-17, doi:https://doi.org/10.1128/genomeA.00829-17 (2017).


